

NBER WORKING PAPER SERIES

A CALL FOR STRUCTURED ETHICS APPENDICES IN SOCIAL SCIENCE PAPERS

Edward Asiedu
Dean Karlan
Monica P. Lambon-Quayefio
Christopher R. Udry

Working Paper 28393
<http://www.nber.org/papers/w28393>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2021, Revised May 2021

We are grateful to Gharad Bryan, Shawn Cole, Andrew Foster, Christina Gravert, Steve Glazerman, Julian Jamison, Heather Lanthorn, Jeff Mosenkis, William Pariente, Kate Vyborny, and especially to Douglas Mackay, for comments and discussions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Edward Asiedu, Dean Karlan, Monica P. Lambon-Quayefio, and Christopher R. Udry. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Call for Structured Ethics Appendices in Social Science Papers
Edward Asiedu, Dean Karlan, Monica P. Lambon-Quayefio, and Christopher R. Udry
NBER Working Paper No. 28393
January 2021, Revised May 2021
JEL No. C81,C83,C9,J01,O12

ABSTRACT

Ethics in social science experimentation and data collection are often discussed but rarely articulated in writing as part of research outputs. Although papers typically reference human subjects research approvals from relevant institutional review boards, most recognize that such boards do not carry out comprehensive ethical assessments. We propose a structured ethics appendix to provide details on the following: policy equipoise, role of the researcher, potential harms to participants and nonparticipants, conflicts of interest, intellectual freedom, feedback to participants, and foreseeable misuse of research results. We discuss each of these, and some of the norms and challenging situations of each. We believe that discussing such issues explicitly in appendices of papers, even if briefly, will serve two purposes: more complete communication of ethics can improve discussions of papers and can clarify and improve the norms themselves.

Edward Asiedu
Department of Finance
University of Ghana
Accra
Ghana
edasiedu@ug.edu.gh

Monica P. Lambon-Quayefio
Econ Dept, University of Ghana
LG BOX 57
IEGON-ACCRA, Ghana
mplambon-quayefio@ug.edu.gh

Dean Karlan
Kellogg Global Hub
Northwestern University
2211 Campus Drive
Evanston, IL 60208
and CEPR
and also NBER
dean.karlan@gmail.com

Christopher R. Udry
Northwestern University
Department of Economics
Weinberg College of Arts and Sciences
2211 Campus Drive #3247
Evanston, IL 60208
and NBER
christopher.udry@northwestern.edu

Introduction

Social science researchers engaged in primary data collection often consider a range of ethical issues during planning, but rarely discuss them in published articles. We believe that building explicit steps for considering and discussing ethical issues can lead to better research, better communications about research, and thus better impact of research as well. We propose a structured appendix to accompany social science papers that report on primary data collection efforts.

We believe that Sen's (1) capability approach provides a useful framework to inform a structured appendix, as it focuses on people's opportunities to set and achieve activities and goals for themselves. This framework, along with Rawls (2), is familiar to most social scientists and applies to both participants and affected non-participants. We acknowledge the diversity of life goals across people, but there is a higher-order common requirement to have the capabilities to set and pursue those goals. Ethical research should be designed to generate socially valuable information to advance these basic capabilities, while protecting the basic interests of participants. This requires research protocols not only be procedurally ethical as per international research standards, but ethical after considering specific contextual factors such as cultural, gender and local institutional norms. London (3) starts from a principle of equal concern, that every participant is the moral equal of all members in the community, and derives several operational criteria that serve as guideposts for this paper and proposal. These include the avoidance of unnecessary risk, special concern for the basic interests of participants, and "social consistency". The latter requires that the sum of incremental risks to participants, minus their direct benefits (call this the net risk) must not be greater than the net risks faced by those in other socially sanctioned activities, like emergency workers. London (3) acknowledges that this calculation is difficult, but "the moral goal of such judgements is clear – the point is to ensure that there is a publicly available justification for the claim that each study participant is treated as the moral equal of every other participant ... and of the community members in whose name research is conducted." These criteria motivate our proposal that authors include a structured ethics appendix in working papers and published online appendices. Such information will highlight the substance of the ethical consideration underlying the study and could also be used in grant applications. We provide a framework below, as a starting point for guidance.

Institutional review boards (IRBs), when present, aim to protect research participants but ultimately examine a narrow set of ethics issues. There are no "ethics review boards" with a mandate over all ethical issues in research. Indeed, a broad range of important ethical issues pertaining to primary data collection research and especially randomized controlled trials (RCTs) and are outside the purview of IRBs. Such issues, often left to self-regulation and peer review during grant applications and publications, deserve more thorough reflective consideration. Active consideration and discussion of these matters can build more consistent norms, improve the philosophical consistency of the norms, and also improve adherence to those norms.

We propose a "Structured Ethics Appendix" for social science researchers engaged in primary data collection (see Appendix). Social science writing style is often fairly terse on methods issues, in particular information relevant for an understanding of ethical issues. Many journals require an IRB approval number and nothing more. This leaves readers to fill in the unstated information with potentially incorrect assumptions. The short-form nature of some public discussions can then exacerbate misunderstandings. The proposed Appendix is not a checklist. Instead, it is a structured, but brief, set of

nine questions which hopefully provide researchers a concise and consolidated platform to spark thoughtful reflection and consideration of ethical issues before a project begins and provide relevant information to readers of completed papers.

The topics discussed are: (1) Policy equipoise, (2) Role of researchers with respect to implementation, (3) Potential harms to participants or nonparticipants from the interventions or policies, (4) Potential harms to research participants or research staff from data collection (e.g., surveying, privacy, data management) or research protocols (e.g., random assignment), (5) Financial and reputational conflicts of interest, (6) Intellectual freedom, (7) Feedback to participants or communities, (8) Foreseeable misuse of research results, and (9) Other ethics issues.

We are writing as economists with experience in primary data collection and design and implementation of randomized controlled trials, not as trained ethicists. Our discussion is undoubtedly, albeit unintentionally, driven by issues that have arisen in our own work. The questions were selected after several rounds of community feedback, from social scientists involved in implementing interventions, and from a wider online audience. We incorporated several suggestions in this version. We intend for the appendix to be a means of prompting discussions too often left in the background, and offer this as a “living document” that we and others can improve through use and feedback. This article accompanies articles on a set of six harmonized randomized controlled trials of community monitoring of common pool resources (CPRs). The trials randomly assigned the introduction of community monitoring to communities (4). Community monitoring is a form of governance that aims to improve CPR management – in these trials, for groundwater, surface water and forests. The community monitoring was typically introduced by non-governmental organizations. We have no reason to believe the researchers failed to adhere to strong ethical practices. However, this kind of collaboration between implementing entities and researchers do make salient many of the points we raise, points that we believe are better off spoken about explicitly. These interventions introduce new governance institutions, and thus ethics questions that pertain to potential harm to both participants and non-participants; the role of the researchers vis a vis the intervention and the implication that has for informed consent; and, the multiple roles of researchers in organizing, reporting, and providing feedback on the interventions themselves needs to explicitly discussed.

1. Policy Equipoise and Scarcity

Is there policy equipoise? That is, is there uncertainty regarding participants’ net benefits from each arm of the study relative to the other arms and to the best possible policy to which participants could have access? If not, ethical randomization requires two conditions related to scarcity: (1) Was there scarcity, i.e., did the inclusion of multiple arms change the expected aggregate value of the programs delivered? (2) Do all ex-ante identifiable participants have equal moral or legal claims to the scarce programs?

Freedman (5) argues that the therapeutic obligation of doctors generates a “clinical equipoise” ethical requirement for medical trials: the expert community must not have certainty that any arm in a trial is better therapeutically than any other arm.² For social science in particular, the word “certainty” can

² There is a robust debate on the meaning and importance of clinical equipoise for ethical medical trials (see 6, 7). Is there a distinction in the role of doctors as researchers and as physicians providing treatment? Or do the ethical principles governing medical therapy apply to clinical research (8)?

render such a requirement toothless: just note that even theories with strong empirical evidence will have some level of uncertainty regarding their applicability in a new context. Furthermore, in many cases one treatment arm clearly dominates another from the perspective of the participants, yet the better treatment arm is not viable as a policy for all, either due to scarcity or other practical or political issues. A study testing the returns to a large cash transfer is a perfect example. Yet evidence from the “better” treatment arm still serves an important social value, even if limited to the generation of abstract knowledge. This renders the concept of clinical equipoise wrong for social science experiments.

Is there a similar obligation for social scientists organizing or participating in RCTs? MacKay (9, 10), building on the same principles as London (2020), argues persuasively that a sufficient requirement for ethical randomization is “policy equipoise”. We have adopted that view here.

Policy equipoise builds on clinical equipoise but considers resource tradeoffs explicitly. Two arms of an RCT are in equipoise when there is meaningful uncertainty about the efficacy of each arm in achieving the relevant outcomes of the study for all participants. Policy equipoise requires that all arms of the study be in equipoise with each other and with the best “proven, morally and practically attainable and sustainable” alternative policy for achieving improvements in the relevant outcomes of the study (9). Each of those words in quotes carries weight. “Proven” implies scientific consensus (and not, for example, adherence to any one methodology for creation of evidence). Clearly, this is not a strictly binary concept, but rather about the degree of certainty of relevant experts and stakeholders, and what is “proven” in one time and place may not be so in another. This also relates to what is meant by “meaningful uncertainty”: naturally, no amount of social science research will render judgements that are perfectly certain; rather, by “meaningful uncertainty” we mean a degree of uncertainty deemed important and likely enough as to make reasonable and informed stakeholders disagree on the optimal policy. “Morally ... attainable” requires that the alternative policy be consistent with individuals’ rights and liberties. “Practically attainable” means that the government or implementing agency has the resources to put the alternative policy into effect. “Sustainable” means that the government or implementing agency could maintain the policy, “given a just system of resource procurement and allocation”(9).

The policy equipoise requirement is a sufficient condition for randomization, but it is quite strict. First, no participant in any arm of an RCT can be predicted to be better or worse off (in an ethically-relevant way, this is, with respect to their capabilities) than he or she would be under any of the other arms of the study. Second, this equipoise must extend to the appropriately defined counterfactual policy, which constitutes whatever policy would have occurred in the absence of the research. Often this counterfactual policy is the current status quo, but that need not be the case. There may be an alternative policy that is moral, practically attainable and for which there is a consensus on its effectiveness (see the hypothetical Progressa case below). Or it could be that the current status quo is not itself sustainable. We emphasize a key distinction that we agree is important, made in MacKay (9), that what matters is not what the actual counterfactual policy is, but rather what it could be.

If policy equipoise is not satisfied, then randomization can be justified if no participant can be predicted to be worse off in any arm of the study than under the counterfactual policy, and if there is scarcity of the resources required for the arms in which participants are better off. A treatment arm may be unambiguously the “best” for recipients but not viable as a systematic policy for all. The simplest example here, relevant in social science, would be large transfer programs. It is unethical to withhold

this policy/treatment unless it is scarce (10).³ Budget constraints can make it impossible to reach everyone with a better costly intervention, in which case it may be permissible to allocate access randomly, if the second scarcity condition is also met, that is, if it is not known a priori who should receive that intervention. If all participants have equal claims to a scarce intervention, then randomization can be ethical. If some participants have stronger claims (for example, if there is local knowledge that the aid would permit some individuals to set and pursue their life goals better – perhaps because they are especially poor and constrained in ways that the aid could relax), then these participants should have priority (11), rather than being randomized out of access.⁴

We highlight one further component of the above definition of policy equipoise: “achieving improvements in the relevant outcomes of the study”. An RCT could, for example, be focused on testing alternatives for reducing child malnutrition. A treatment arm could be cash transfers. If we were to “know” that cash transfers create benefits beyond reducing child malnutrition (a reasonable conjecture, albeit one with uncertainty regarding the types and magnitudes of impacts), should the non-child malnutrition outcomes be considered when examining if any treatment arm is viably “better” than all of the others? Doing so would require a full-blown multi-dimensional welfare analysis, weighing all ethically-relevant outcomes. We argue that it is typically sufficient to constrain the comparison of treatment arms (and counterfactual policy) to the problem at hand (e.g., paths to improving the primary outcomes of the study). Realistically, anything that dominated in all dimensions with certainty would likely not be “practically attainable.” We do see merit in this discussion, though.

Randomization into different treatment arms, therefore, is unethical if two conditions are met. First there is not policy equipoise. And second, either the preferred arm does not require resources that are scarce or all participants do not have equal claim to those scarce resources. Some examples can clarify the requirements of policy equipoise and scarcity.

Karlan et al (13) implements an RCT that compares a treatment arm in which farmers are offered cash grants of approximately \$400 to a control group. This RCT violates policy equipoise, because there is no meaningful uncertainty among experts that receiving an offer of a cash grant is better for the recipient than not receiving such an offer. However, randomization remains ethical because of the two principles of scarcity. Providing cash grants to farmers predictably improves their welfare, so the cash grant arm is better than the counterfactual policy of no cash grants. However, there is no consensus that providing cash grants to farmers as a policy is practically attainable and sustainable, given the cost of that policy. This meets the first requirement for scarcity. Nor is there a consensus about which farm households have stronger claims over receipt of cash grants. This meets the second scarcity requirement, and therefore randomization of access to the grants is acceptable.

Similarly Glewwe et al. (14) violates policy equipoise – there is no uncertainty that eyeglasses benefit those whose eyesight can be corrected by them. However, there is no consensus that eyeglass provision

³ A salient example of an instance where there is *not* scarcity is the policy decision about how a given program of cash grants for child welfare is to be transferred to a married couple. Should the grant be given directly to one or the other or both together? Or divided between the two? If there was a consensus among experts that providing child care grants to the female of a couple improved child welfare more than providing the grant to the man, without other adverse consequences, then randomization to provide grants to men would not have a scarcity justification (and also would violate policy equipoise).

⁴ The growing selective trials literature provides a promising path forward on capturing local information (12).

to school children is a cost-effective policy for improving school achievement. Therefore, it was not clear a priori that there was an alternative, sustainable policy that dominates the control arm, satisfying the first scarcity requirement. Randomization occurred at the township level, and there was no a priori knowledge of which township would benefit most from the intervention, satisfying the second scarcity requirement.

Policy equipoise would be violated if there is an expert consensus that an arm of a trial is dominated by another policy that is proven, attainable and sustainable. A hypothetical but realistic example of a violation of policy equipoise would occur if a government with sufficient resources for a national program agreed to withhold a Progres-a-inspired conditional (on school enrollment) cash transfer (CCT) from some households for a pure control group in an RCT (10).⁵ In this hypothetical case, the first scarcity requirement also fails, because the government is presumed to have sufficient resources to make the CCT available to more households, so this randomization is unethical.

Consider a potential research project studying the effect of mobile money transfers on food security during a drought with a control group receiving nothing and a treatment group receiving the transfers. While policy equipoise is violated because the transfers can be expected to improve food security, the first scarcity requirement is met due to the budget constraint of the project. However, if the researchers know a priori through, say, remote sensing data, which households' farms were particularly affected by the drought, the second scarcity requirement may be violated. The most affected farmers should be prioritized for transfers, rather than randomized into treatment or control.⁶ Thus, such a proposed randomization raises an ethical red flag.

2. Researcher Roles with Respect to Implementation

Are researchers “active” researchers, i.e. did the researchers have direct decision making power over whether and how to implement the program? If YES, what was the disclosure to participants and informed consent process for participation in the program? Providing IRB approval details may be sufficient but further clarification of any important issues should be discussed here. If NO, i.e., implementation was separate, explain the separation.

In social science, the role of the researcher is quite varied, and this has important implications for what may constitute ethical research. On one end of the spectrum, the researcher is merely the evaluator, with neither influence over nor responsibility for any of the interventions. In such an instance, understanding the ethics of the intervention may even be the motivation for the research. Then with more evidence on deleterious effects of the intervention, perhaps such policies can be altered or eliminated.⁷

On the other end of the spectrum, the researcher is the implementer (“active” in our classification). In such cases, the researcher for example secured funding, hired staff, decided the design of the

⁵ Policy equipoise could be restored if the control group were replaced by an experimental arm expected to be as effective or better as the CCT.

⁶ Naturally “most affected” is undoubtedly mismeasured and not known with uncertainty; as such this could be done probabilistically, in which individuals' likelihood of assignment to treatment is a function of estimated prioritization. This maintains the ability to test impact throughout while still adhering to this ethical principle.

⁷ See Glennerster and Powers (15) for a robust discussion of this distinction and its implications, and some further helpful examples.

intervention, and directly implemented it. In this case, it goes without saying that the researcher is responsible for the ethics of the intervention itself. The researcher is changing the world not merely through the dissemination of results of the research but deliberately through the research process itself.

Even when the researcher is “active”, however, it may still be “ethical” to implement arms which raise valid ethical concerns. But certain criteria ought to be met. First, there must not be a consensus among the relevant expert community about the outcomes that raise the ethical concerns (and this may be the aim of the study). Second, the policy or intervention must be commonplace enough that assessing the intervention is of high social value. In these instances, the onus is on the researcher to establish “policy equipoise”, or explain why scarcity limits the extension of expected beneficial arms to the entire population as discussed in Section 1 and MacKay (10).

Two recent examples in social science highlight the importance of this distinction, and also point to the complicated nature of defining the researcher’s role in some instances. In Bryan, Choi and Karlan (16), researchers collaborated with a nonprofit organization in the Philippines which ran a four-month program that included various secular activities (e.g., savings groups, livelihood promotion, health education) alongside weekly meetings with a pastor that covered an evangelical Protestant curriculum. To understand the impact of the Protestant curriculum on individuals and their households, the nonprofit organization agreed to randomize across villages whether or not the Protestant curriculum was included in their program. The research did not lead to an increase or decrease in the quantity of sessions conducted, nor the number of people reached, by the program. Of note, there is no consensus on the impact of such programs, and they are commonplace in the world. Thus, we argue that there is positive social value in evaluating claims of impact or harm, and no ethical quandary for the researchers.

In Kenya, researchers from universities as well as the World Bank collaborated with the Nairobi Water and Sewerage Company, a Kenyan utility company regulated by the government of Kenya. The researchers engaged with the utility company over several years, working broadly and collaboratively in order to increase access to water and sanitation services. Coville et al. (17) reports on this collaboration, specifically on an experiment in which one of the treatment arms threatened tenants with having their water access shut off if the water bill was not paid by their landlord. While one could debate the ethics of that intervention arm as a policy, we would argue that if the researchers unambiguously have no “active” role then such a discussion is important and worthwhile but should be discussed as an ethical consideration for the utility company, not the researcher. Suppose counterfactually that the utility company was already shutting off water of tenants whose landlords had not paid and researchers worked with the utility to organize an RCT that shielded some from being shut off. This version of the trial is identical to that described above. However, because the status quo policy potentially infringes on important human rights (and in fact, cutting off utilities due to non-payment is a common policy around the world), the ethical issues may make this more important to study. Ultimately, this study also highlights potential ambiguity on the binary categorization of “active.” If a researcher suggests a treatment arm, for example, does this make them responsible for its ethics, even if the researcher has no actual power over the decisions? Putting ethical onus on the researcher could be construed as ignoring (and thus disrespecting) the autonomy and agency of the local policymakers. However, researchers should acknowledge their role in influencing policies and decisions, particularly in long-standing and complicated collaborations, and thus take ethical responsibility for any such influence. We argue that more transparency on this role would be fruitful for social science, so that norms can be better developed on this question.

Extreme examples, however, render this distinction meaningless. For example, research ethicists share a common consensus that medical data from experiments conducted on prisoners in Nazi concentration camps cannot now be used to answer medical research questions. But, even if one believes the Vietnam War was unethical, using the lottery draft to study the impact of military service on later earnings is easily defensible as ethical. Perhaps one key distinction is the purpose of the original implementation, because of the incentives created for researchers. If the original implementation was being done for the sake of research (albeit unconscionable and unethical), then allowing future researchers to use such data creates perverse incentives to implement unethical research. However, when the original implementation was “natural” (e.g., the Vietnam lottery example above, or a government-run public lottery for secondary school scholarship (18), or reservation of village leader slots for women (19)), then we argue that use of such variation is valid ethically even if there are objections to the underlying policy.

3. Potential harms to participants or nonparticipants from the interventions or policies

Does the intervention, policy or product being studied pose potential harm to participants or non-participants? Related, are participants or likely affected non-participants particularly vulnerable? Also related, are participants’ access to future services or policies changed because of participation in the study? If yes to any of the above, what is being done to mitigate such risks?

We highlight two broad issues. First, despite best efforts, IRBs are unable to fully oversee all aspects of potential harms to participants, and local knowledge in particular is invaluable and at the mercy of self-regulation. Second, potential harms to nonparticipants is itself a topic of high interest to researchers on substantive levels, but thus also carries important ethical considerations.

Ethical guidelines for research on human subjects are primarily concerned with protecting study participants (Eyal et al. (20)). Protocols typically expect researchers to highlight the benefits and risks of participation, confidentiality protocols, compensation, and instructions for withdrawing from the study at any stage. While in theory IRBs strive to consider all such issues, in reality IRBs, particularly ones from a different culture, are not always informed enough to raise or adjudicate on them (due to omission by researchers or lack of knowledge of context and practices). Ultimately the responsibility of the researcher extends beyond mere IRB approval. We emphasize, as well, that “harms” extend far beyond mere monetary harm, in particular for example for marginalized populations, such as women in many cases, that may be engaged in non-wage labor.

Cultural norms and nuances of the implementation may pose potential harms to participants that are beyond the viable purview of the IRB. For example, we have experienced in Ghana and Nigeria cultural norms or misinterpretations of implementation activities that have led participants to be perceived as being “superior” or to suffer stigma or discrimination, which were mitigated by providing additional information or community discussions. If not addressed, such perceptions can boost or harm the reputation of participating households which may have implications for their social and perhaps economic lives. Elected community leaders in participating communities may suffer abuse and intense competition from political rivals, as interventions are sometimes perceived vehicles for local leaders to maintain political power. When deemed plausible projects ought to integrate mechanisms to deal with such issues.

While informed consent and IRBs are primarily concerned with study participants, research may also pose harm to nonparticipants. The National Bioethics Advisory Commission of the US asserts that

regulatory oversight for research with human subjects extends beyond the protection of individual research participants to include the protection of social groups (21). The potential effects on nonparticipants become especially salient when researchers are actively engaged in interventions. It is, therefore, important to evaluate the scope and intensity of risks that are likely to impact nonparticipants. In some cases, community-level consent may be appropriate (see section 4).

We highlight two examples in which nonparticipants are affected and discuss the ethical implications.⁸ Ashraf et al. (23) conducts an experiment in Zambia on women's bargaining power and fertility. In one arm women were given access to contraceptives alone and in the other with their husbands. Since both arms of the trial have implications for a nonparticipating husband in terms of child-bearing outcomes, should there be household-level consent for both arms? In this case, no, because obtaining such household level consent could trigger domestic violence. Women's control of their reproductive decisions is a basic human right, as is their freedom from domestic violence. A woman's nonparticipating husband may be affected, but her human rights must take precedence.

Second, incentivizing students to take part in an antiauthoritarian protest poses a direct risk to participants but also poses risks to non-participants engaged in the protest (see 24). Thus, if students may reasonably be expected to be rowdier in a protest than non-students, incentivizing more student participation in a protest (may?) increase the risk for the non-student nonparticipants of the intervention. Inasmuch as it is important to minimize the risk for the study participants, it is in this case even more critical that the actions of participants do not increase the risk for nonparticipants.

Consideration of risks and benefits for non-participants helps evaluate the ethics and establish appropriate mitigation strategies. Naturally, however, indirect effects are easy to posit qualitatively but much harder to confidently predict quantitatively. And, indeed, studying indirect effects is itself a robust research agenda (e.g., 25). Thus, we do not argue that the mere posing of a possible indirect effect should render a study unethical. Rather we suggest considering such issues, and as we learn more about them from ongoing research, we hope this ethical conundrum can become more evidence-based.

We acknowledge that there are unresolved issues in expanding ethical considerations to groups, which require further discussion and consensus-building. Research risks to communities and nonparticipants suggest that ethical protocols focused on individual research participant protection will be insufficient. Ethical guidance for research in social science in general and RCTs specifically could extend protection beyond mere individuals to social groups i.e., family members, community members, etc. (21). Group implications of identified individual-level harm can serve as a starting point to identify risk to nonparticipants and should be properly assessed and addressed in an ethics review process. Working directly with community representatives may prove fruitful to discuss informed consent and develop methods that minimize potential group harms.

4. Potential Harms from Data Collection or Research Protocols

Are data collection and/or research procedures adherent to privacy, confidentiality, risk-management, and informed consent protocols with regard to human subjects? Are they respectful of

⁸ McDermott and Hatemi (22) provides additional examples.

community norms, e.g., community consent not merely individual consent, when appropriate?⁹ Are there potential harms to research staff from conducting the data collection that are beyond “normal” risks?

The fundamental principle of ethical conduct of human subjects’ research centers around minimizing the risk of harm and respecting agency (i.e., securing appropriate informed consent).

Informed Consent: Data Collection

Genuine consent is present only when participants have the capacity to fully understand and process the shared information (26). Processes for obtaining informed consent differ widely, and stated protocols and actual practice may differ. Field staff may treat consent as a mere formality where the main task is to get the participant’s approval, thus leading field workers to not fully engage participants. If interviewer and respondent differ on age, education or status levels, undesirable expectations may ensue due to perceived power or influence. Electronic data collection offers paths to oversight, e.g., with time stamps on informed consent and survey components. However, perfect monitoring is not possible; staff training is critical. Furthermore, gifts, even minor ones such as soap, may have the unintended consequence of distracting from the importance of the informed consent; compensating someone for their time, on the other hand, is also an important principle to respect.

Enumerators are often encouraged to probe during their interviews to obtain more accurate responses. While probing can be useful, it can also create discomfort, be considered intrusive, and expose vulnerabilities of the participant. There is a fine line between probing and harassment.

Moreover, although consent forms provide participants with researcher contact details, they may be less likely to contact the researchers when they experience any privacy violations because of telecommunication charges or perhaps perceived language or status barrier.

Informed Consent: Randomization

Informed consent on the randomization process itself is often not discussed. In medical trials, partly because the medical professional has a “standard of care” therapeutics to which each patient has a clear right, informed consent on the randomization process is essential to having agency over one’s own health decisions. However, in social science, the intervention is often implemented independently from the researchers (see Section 3), and “informed consent” is effectively implied via voluntary participation in an intervention (e.g., if a lender randomly offers people different loans, their decision to borrow is not forced by the lender nor by the researcher, and consent is implied by their decision to apply and/or sign a contract for a loan). Furthermore, in social science scarcity often limits the total number of individuals that can receive a certain intervention, and the randomization is being done not for research purposes but rather to allocate a scarce resource fairly and with minimal risk of corruption (e.g., see 18).

⁹ Example of sub-questions to consider as part of the broad question: Are there any risks that could ensue because of the data collection process or storage, e.g. discomfort to being asked certain questions or breach of confidentiality? If so, what are the mitigation strategies? Are there costs to the participant for the data collection process, such as their time, and if so, what is the strategy or rationale for offsetting this cost? Because these are all issues covered by most IRB processes, a sufficient explanation for a “yes” response may be to provide the IRB approval numbers for all IRBs that have approved the project. However, if there are particular issues that are important to discuss, please do so here.

Hawthorne and John Henry effects are often discussed as threats to internal validity, but they each generate ethical issues as well (and not merely by reducing the social value of the research). A Hawthorne effect is generated when individuals, because they know they are in a treatment group, behave differently. A John Henry effect is the analog for a control group. A problem ensues if there is an “unlucky stigma” among individuals leading them to harbor negative psychological effects (for control group participants). The analog for a treatment group is potentially more nuanced. If, for example, a treatment participant receives a cash transfer, observing jealousy, demands for sharing, or other such effects, are not necessarily a risk to internal validity, but rather perhaps exactly the kind of changes in behavior that the researcher (and policymaker) seek to learn about. The question for ethics is whether the randomization process itself generated such effects over what the transfer itself generates. This is difficult to know, empirically, but is a risk nonetheless. When such a risk is deemed likely, we suggest a discussion.

Informed Consent: Individual vs Community

Social norms in many developing countries, particularly in sub-Saharan Africa, raise important issues regarding individual versus group level consent. For example, often local norms mandate that visitors to communities first meet with chiefs and community leaders, explain their reasons for visiting the community, and in some cases provide customary gifts, such as presenting kola nuts or drinks. This level of community consent improves the security of field enumerators and researchers. This also may be necessary etiquette for the sake of future researchers, so that future enumerators are not greeted with mistrust. Community consent also could be critical for managing local security risks that may arise (and due to timing, be outside the practical purview of the IRB process).

Aside from issues relating to obtaining individual versus community consent, there are other secondary level ethical concerns that projects ought to consider in the process of obtaining community or individual level consent. In some settings research projects may inadvertently contribute to creating an expectant culture of participants or facilitate the corrupt behavior of local community and political leaders. For example, community leaders whose assistance may be required during the community entry or obtaining community consent for an intervention may take advantage of the project to advance their own agenda or illegally take bribes from community members in exchange of providing some service that the project may be providing for free.

Local IRBs can prove particularly useful for these issues, as they may have better information about the local expectations and norms.

Privacy, Confidentiality and Sensitive Information

Once data are collected, the responsibility of ensuring confidentiality of the information gathered rests on the researcher. Researchers are expected to not betray the trust of participants with respect to data management, storage, and protection of privacy. Failure to ensure the confidentiality of the participants may have serious consequences. For example, revelation of information on the health status, particularly sexually transmitted diseases such as HIV, or sexual behavior of a participant can cause harm to participants. Similarly, financial data could be used by a financial institution in ways contrary to a client’s self-interest. Biomarker data, which open the possibility of disclosure of genetic information, raise the stakes further. Confidentiality is, therefore, key to safeguarding and reducing the risk of potential stigmatization of participants, especially for a vulnerable population. In some cases, the

potential harm of a lax data management system may extend beyond the individual participant to his or her extended family.

For sensitive questions (e.g., sex, health, politics, religion, etc.), discussing the specific survey methods may help both for explaining the adherence to ethical protocols but also to guide future researchers looking to ask similar questions.

Potential Harms to Field Staff

Ironically, risks to field staff are not a consideration for most (if not all) IRBs, since they are not research participants. Yet clearly field staff are often unsung heroes of the research effort; without high quality data, a randomized controlled trial is akin to an elaborate birthday party without guests.¹⁰ But field work does create risks to staff. Risks range from “normal” such as road accidents or verbal abuse to extreme such as sexual harassment or political violence. Pandemics of course pose even further risks to communities, in which traveling staff may unintentionally worsen the spread of a disease particularly in situations like COVID-19 where carriers are often asymptomatic.

5. Financial and reputational conflicts of interest

Do any of the researchers have financial conflicts of interest with regard to the results of the research? Do any of the researchers have potential reputational conflicts of interest?

Financial conflicts of interest reporting policies are similar but distinct across universities, multilateral organizations, and research conduits such as the National Bureau of Economic Research or the American Political Science Association. It is important to report the ex-ante possibility of a conflict of interest, so that the research community can judge for itself whether the interpretation of the empirical results was biased in favor of the possibly competing interest. These rules become vaguer when the income source and the data source differ but are from the same industry. If earning consulting income from Bank A, and writing a paper with data from Bank B, does the Bank A consulting income need to be reported in the research about Bank B? Ultimately researcher judgement is required to make such decisions; the goal of the structured appendix is to report potentially important issues without being a burden.

Reputational conflicts of interest are not addressed by any university, multilateral or research organization rules that we have seen. But for many researchers, idea promotion is a vital self-interest (whether due to altruism, reputation, or future financial remuneration). While money is an interest that is easily defined, and typically traceable via contracts and payment records, reputational self-interest is person-specific, and difficult to define and trace. We thus define a reputational conflict of interest with respect to a particular paper quite broadly: when prior writing or advocacy could be contradicted by specific results in the new paper, and such contradiction would pose reputational risks to the author. Taken to an extreme, such disclosures could be exhausting (and thus then ignored). We do not recommend that. We make a key assumption: *academically*, there is no negative consequence to publishing work that contradicts one’s earlier work (indeed, the effect could be the opposite, to increase credibility). Rather, here, we are referring to reputation outside of academia, where a public presence

¹⁰ Analogy courtesy of Christina Gravert, with the follow-on explanation: “Tons of work to set up and then makes you cry out of disappointment.” An alternative and subtler analogy, courtesy of Andrew Foster: “A compass in Antarctica”.

has been built around prior research findings or arguments. We provide two examples from our own research.

First, for Karlan, his paper “Tying Odysseus to the Mast” (27) evaluates the take-up and impact of a commitment savings product in the Philippines. He subsequently published a paper on commitment contracts for smoking cessation. Both find generally positive impacts from the commitment devices. Subsequent to the above paper, Karlan started a for-profit company stickK.com that helps individuals write commitment contracts to change some future behavior (typically weight loss, exercise or smoking). This company has been written about frequently in the media. Karlan owned equity when he co-founded stickK.com, but has subsequently transferred all of his equity to charitable structures and he receives no financial remuneration from stickK.com. As such he no longer has a financial conflict of interest that is reportable per standard university rules. But he does have a reputational conflict of interest on the topic of commitment contracts, as his work and stickK are often cited in the media as an example of successful “nudges”. A disclosure would be appropriate for Karlan on future work related to committed contracts.

For Udry, his paper “Gender, Agricultural Production, and the Theory of the Household” (28), finds that households do not achieve pareto efficiency (yet much empirical work using household data starts with such an assumption and never looks back). If Udry were to have a new paper in which he finds pareto efficiency, this would conflict with his prior finding but would not constitute in our opinion a reputational conflict of interest. Even though his earlier work is commonly cited for testing for efficiency within the household, Udry has no non-academic reputation that revolves around that finding (and also, for the record, the other three of us are certain Udry would be quite happy to find such evidence, and then would undoubtedly try to concoct some neoclassical economic theory as to why that context yielded such a result, but other contexts did not).

Lastly, coauthorship with non-conflicted researchers and registration of RCTs and pre-analysis plans both can serve to mitigate reputation conflicts of interest issues because they both tie the hands of the research to some extent. Neither of course is dispositive: the choice of specification and the choice of outcomes (timing, measurement approach, and selection of proxies) could all be biased. And naturally the way results are interpreted (even with refereeing and editing at journals) can be biased. Thus, reporting such conflicts, and what steps if any were taken to mitigate, could be useful.

6. Intellectual freedom

Were there any contractual limitations on the ability of the researchers to report the results of the study? If so, what were those restrictions, and who were they from?

While reporting of conflicts of interest makes potential biases transparent, parties with vested interest may have other means of controlling research output and thus creating bias. Specifically, researchers and collaborators (here we define collaborators as either implementing parties, sources of data, or funders) often sign contracts regarding funding, intervention plans, or data rights. If such contracts infringe on the intellectual freedom of the researcher, this poses harm to the credibility of the research. Failure to disclose such conditions could reduce the social value of the research, constituting an ethical transgression. As such, we have included an explicit question on this in the appendix.

The most egregious infraction is the simplest: a contract which provides an external party (funder or implementer) unilateral power to prevent publication of the results. This should be unacceptable to any researcher interested in putting forth credible research, but at a minimum if such were the rights of the partner this ought to be disclosed in any and all publications (as in 29).

We consider four restrictions to be benign but still worthy of reporting for the sake of completeness: (1) permission required to report the name of the collaborating institution, (2) a comment commitment: collaborating institution can provide comments on the research output and the researchers agree to consider these comments to the best of their judgement (but are not committed to incorporating them), (3) a timing commitment: the collaborating institution has a right to see the results before others, with a specified period of time, and (4) proprietary intellectual property: the collaborating institution has developed a proprietary technology (perhaps an algorithm, for example) that is relevant for the research, but can only be described broadly without revealing its inner workings (such restrictions limit the scientific value of a publication, which could pose an ethical transgression). We do not assert this to be a complete list of all acceptable restrictions, but we are unaware of any others that we consider acceptable.

A few details regarding researcher independence are important to note:

- Timing matters: Intellectual freedom must be granted in advance. A clause in an agreement which states that this will be decided later is a farce to the concept of researcher independence.¹¹
- Partial set of coauthors: If a subset of the coauthors on an academic paper are conflicted or lack independence (e.g., most commonly, because they are employed by the funder or the implementing organization), then this section of the appendix ought to make clear that the independent researchers had full rights to publish on their own without the conflicted coauthor.

7. Feedback to participants or communities

Is there a plan for providing feedback on research results to participants or communities? If yes, what is the plan? If not, why not?

Informed consent ultimately is about respect for the agency of an individual as well as their privacy and property rights to their information. But often attention to this focuses on the informed consent process. Recently more attention, but still too little, has been placed on engaging with research participants, local government, and communities after the research is complete (30).

To further the principles of respect for persons, beneficence, and justice, researchers should consider providing feedback to research participants and communities (22, 31). Providing such feedback reduces the likelihood that the participant may feel exploited by the researcher. Public awareness of research results also may be a stepping stone towards research uptake to improve the lives of poor communities. In addition, providing feedback also improves the willingness of communities to participate in future research. Note that in many cases, key gatekeepers (community leaders) may have been part of

¹¹ In our collective experiences, we have only encountered such a request once (specifically, the assertion was “to decide researcher independence later, after results are in”). The individual arguing that this met a reasonable standard of researcher independence led a large, corporate “philanthropy” effort, and subsequently left their post at the corporation to take a high-ranking position with the Trump administration.

community-level consent (see above), and as such the feedback may be appropriate to such gatekeepers (when not viable to all participants).

Despite the ideal of providing such feedback, there are several practical challenges that often render this aspiration unwise. We highlight three: budget, finding people, literacy/knowledge necessary to understand the research.

Budget

Follow-up visits to all communities (and finding people) can be costly and in some cases these costs can outweigh the benefits of directly providing feedback. However, with cellphone data and internet access expanding rapidly, communicating results of the research to participants could be viable, e.g., posting of information on a website and sending individuals text messages, or sending pre-recorded video or audio via text message.

Literacy/knowledge necessary to understand the research

Depending on the context and the research, study participants may struggle to understand research findings. Poor understanding of research findings can generate emotional distress for research participants and affect community involvement in future studies. Therefore, providing feedback to participants should not be done solely for the purpose of ticking a box, but with genuine engagement. Furthermore, any study with nuanced results may get misused intentionally or unintentionally. Researchers must examine their context so that proper measures can be taken to enhance understanding of their results. We propose that the offer of feedback to participants be included as part of the informed consent processes for social science research, although IRBs may raise concerns if this offer created undue influence to participate (perhaps due to unrealistically optimistic beliefs about the value of that feedback). As also argued in the medical literature, this process of providing feedback must be designed to provide clear opportunity for the participant to decline receipt of results (see 32).

For policy-oriented research, it may be effective to provide feedback at a much higher level than at the individual level. Policy prescriptions often are made at the community, district, or national level. Researchers should consider the level at which to provide feedback to be relevant for policy.

Providing feedback to participants may itself affect behavior. Before implementing such a process, the possible adverse consequences of providing feedback should be considered. Given the complexity of some research, a risk is that feedback is conveyed but then misconstrued, leading to detrimental actions. This concern reinforces the importance of careful communication and thoughtful consideration of the way the results may be interpreted. And, of course, this is a question for research itself, does provision of such feedback change later behavior?

8. Foreseeable misuse of research results

Is there a foreseeable and plausible risk that the results of the research will be misused and/or deliberately misinterpreted by interested parties to the detriment of other interested parties?

In research settings characterized by strong imbalances of power between interested parties, there may be foreseeable, plausible risks that a powerful party might use findings from the research in ways that will harm participants or nonparticipants. Research might reveal a vulnerability of a sub-population that

can be exploited for the gain of a more powerful party. If this is the case, the researcher has an ethical obligation to mitigate these risks.

An example might be an innovation in microcredit that research shows is overall profitable for lender and borrower on average. However, consider heterogeneity. First, suppose that the microcredit innovation harms a vulnerable and identifiable minority of borrowers. Second, suppose that while the innovation leads to an increase in borrower income, it has other psychological or social costs to borrowers that outweigh the income gain for at least some participants. In addition, suppose there is a foreseeable and plausible risk that for-profit lenders could misuse the “average” research results to advocate for expansion of the innovation without mitigating the deleterious impacts. This observation calls for the research to take steps to address the possibility of such misuse of the research results by considering such risks ex-ante and incorporating a measurement and communication strategy to ensure a complete and balanced reporting of the results.

In a context of a repressive or non-democratic government, additional care is required. Location-specific research to improve digital identification services in such a context, for example, raises the possibility that the results could be used for repressive purposes and researchers should consider this possibility when considering and designing their work.

We use the phrase “foreseeable and plausible” to limit the scope of this question. Research findings can be misused in many ways that should not be considered to be the responsibility of the researcher (companies making use of the discovery of a novel behavioral pattern to sell unneeded products; the police force of a repressive government using techniques discovered in research elsewhere on making bureaucracy function more efficiently). The ethical responsibility of the researcher is to consider harmful uses of their work that are predictable and reasonably tied to the research context.

9. Conclusion

While much has been written on ethics of research, we perceive there to be a large empirical research gap. Broadly we see two categories of empirical research needed: documenting contextual factors that render a particular ethical concern critical or negligible, and learning field research methods to improve adherence to stated intentions.

Examples of the first category include the following: Under what circumstances does random assignment (whether to treatment or control) generate stigma that harms participants? Does promising the control group a later service change their behavior now, during the observation period of the study?

Examples of the second category include the following: How much of an informed consent is understood by participants, and what wording and content changes lead to higher or lower informed consent comprehension rates (e.g., see 33, 34)? What survey methods and staff trainings are most effective for reducing harm from data collection, e.g. from adverse emotional reactions to sensitive questions? When does conducting surveys change later behavior, i.e., a mere measurement effect (e.g., see 35).

We hope that the proposed structured ethics appendix will accompany papers, and spark further consideration of ethical issues in field research. We also would support inclusion of these questions, or a subset therein, in grant applications. Furthermore, the anticipation of writing such appendices may shape design decisions. We have aimed for brevity, understanding that anything too burdensome will not be adopted. In a public comment period, we received several useful suggestions to include further

questions. We adopted several, but not all. For example, some suggested *ex post* discussion of actual harms. We decided not to include this because those are for the most part covered by IRB protocols, and we also believe the ethical decisions ought to be adjudicated blind to results; otherwise “no harm happened” could be misconstrued as evidence of an ethical design. This risks, however, that failure to discuss realized harms may thwart improvements in ex-ante reasoning by researchers using similar methods. This may be an area that shifts, if our proposed guidelines are adopted and adapted. We also did not extend the reporting of potential harms to be reporting of potential welfare change. Although we admire the aspiration, we fear that such an inclusion would inevitably require considerable work and untenable assumptions in order to assert. Furthermore, loss aversion in both law and ethics is commonplace. Lastly, we did not include a call to discuss the social value of the research with respect to potential policy. Although this is an important consideration and articulating the policy implications may improve the research-to-policy nexus, we believe this would be too burdensome if addressed more robustly than merely regurgitating points made in most paper introductions. We did not receive any suggestions for items to remove, although a common suggestion was to make this shorter to increase participation and impact.

We cannot imagine these nine questions, as posed, being perfect, and we hope the by-product of this proposal is a living document which gets updated over time as usage is observed, questions improved, and norms change. We are posting the proposed structure and seeking further comments on the Global Poverty Research Lab website at Northwestern University, and will also be promoting this to peer institutions for cross-posting. We will also monitor actual usage of structured appendices, to identify any improvements authors make that can be incorporated into the guidelines.

Research ethics is ultimately self-regulated, and our goal is to create a norm of discussing these issues explicitly, both to strengthen and improve the norms and research practices and also to improve the discussion of these issues. To do so we believe that all cases ought to be discussed, not just the ones that raise questions. We hope that a norm of inclusion of a structured ethics appendix can help the research community advance.

References

1. A. Sen, *Commodities and capabilities* (North-Holland ; Sole distributors for the U.S.A. and Canada, Elsevier Science Pub. Co, 1985).
2. J. Rawls, "Social unity and primary goods" in *Utilitarianism and Beyond*, A. Sen, B. Williams, Eds. (Cambridge University Press, 1982), pp. 159–186.
3. A. J. London, "Equipose: Integrating Social Value and Equal Respect in Research with Humans" in *The Oxford Handbook of Research Ethics*, A. S. Iltis, D. MacKay, Eds. (Oxford University Press, 2020) <https://doi.org/10.1093/oxfordhb/9780190947750.013.13> (March 15, 2021).
4. T. Slough, D. Rubenson, R. Levy, et al, Adoption of Community Monitoring Improves Common Pool Resource Management Across Contexts (2021).
5. B. Freedman, Equipose and the Ethics of Clinical Research. *N. Engl. J. Med.* **317**, 141–145 (1987).
6. F. Gifford, So-Called "Clinical Equipose" and the Argument from Design. *J. Med. Philos.* **32**, 135–150 (2007).
7. W. Chiong, The Real Problem with Equipose. *Am. J. Bioeth.* **6**, 37–47 (2006).
8. F. G. Miller, H. Brody, Clinical equipose and the incoherence of research ethics. *J. Med. Philos.* **32**, 151–165 (2007).
9. D. MacKay, The ethics of public policy RCTs: The principle of policy equipose. *Bioethics* **32**, 59–67 (2018).
10. D. MacKay, Government Policy Experiments and the Ethics of Randomization. *Philos. Public Aff.* **48**, 319–352 (2020).
11. C. B. Barrett, M. R. Carter, The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections. *Appl. Econ. Perspect. Policy* **32**, 515–548 (2010).
12. S. Chassang, G. Padró I Miguel, E. Snowberg, Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments. *Am. Econ. Rev.* **102**, 1279–1309 (2012).
13. D. Karlan, R. Osei, I. Osei-Akoto, C. Udry, Agricultural Decisions after Relaxing Credit and Risk Constraints. *Q. J. Econ.* **129**, 597–652 (2014).
14. P. Glewwe, A. Park, M. Zhao, A better vision for development: Eyeglasses and academic performance in rural primary schools in China. *J. Dev. Econ.* **122**, 170–182 (2016).
15. R. Glennerster, S. Powers, "Balancing Risk and Benefit: Ethical Tradeoffs in Running Randomized Evaluations" in *The Oxford Handbook of Professional Economic Ethics*, G. F. DeMartino, D. McCloskey, Eds. (Oxford University Press, 2016), pp. 366–401.
16. G. Bryan, J. J. Choi, D. Karlan, Randomizing Religion: the Impact of Protestant Evangelism on Economic Outcomes. *Q. J. Econ.*, qjaa023 (2020).

17. A. Coville, S. Galiani, P. J. Gertler, S. Yoshida, “Enforcing Payment for Water and Sanitation Services in Nairobi’s Slums” (Social Science Research Network, 2020) (December 31, 2020).
18. J. Angrist, E. Bettinger, E. Bloom, E. King, M. Kremer, “Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment” (National Bureau of Economic Research, 2001) (April 18, 2016).
19. R. Chattopadhyay, E. Duflo, Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica* **72**, 1409–1443 (2004).
20. N. Eyal, M. Lipsitch, T. Bärnighausen, D. Wikler, Opinion: Risk to study nonparticipants: A procedural approach. *Proc. Natl. Acad. Sci.* **115**, 8051–8053 (2018).
21. R. R. Sharp, M. W. Foster, Community involvement in the ethical review of genetic research: lessons from American Indian and Alaska Native populations. *Environ. Health Perspect.* **110**, 145–148 (2002).
22. R. McDermott, P. K. Hatemi, Ethics in field experimentation: A call to establish new standards to protect the public from unwanted manipulation and real harms. *Proc. Natl. Acad. Sci.* **117**, 30014–30021 (2020).
23. N. Ashraf, E. Field, J. Lee, Household Bargaining and Excess Fertility: An Experimental Study in Zambia. *Am. Econ. Rev.* **104**, 2210–2237 (2014).
24. L. Bursztyn, D. Cantoni, D. Y. Yang, N. Yuchtman, Y. J. Zhang, Persistent Political Engagement: Social Interactions and the Dynamics of Protest Movements. *Am. Econ. Rev. Insights* (2021) <https://doi.org/10.1257/aeri.20200261> (January 4, 2021).
25. E. Miguel, M. Kremer, Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica* **72**, 159–217 (2004).
26. S. E. Hewlett, Is consent to participate in research voluntary? *Arthritis Rheum.* **9**, 400–404 (1996).
27. N. Ashraf, D. Karlan, W. Yin, Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines. *Q. J. Econ.* **121**, 635–672 (2006).
28. C. Udry, Gender, Agricultural Production, and the Theory of the Household. *J. Polit. Econ.* **104**, 1010–1046 (1996).
29. EGAP, “EGAP Research Principles” (2011) (March 15, 2021).
30. S. Friedlander, *et al.*, Sharing Research Results with Participants: An Ethical Discussion. *MIT Poverty Action Lab Manuscr.* (2021).
31. C. V. Fernandez, E. Kodish, C. Weijer, Informing study participants of research results: an ethical imperative. *IRB* **25**, 12–19 (2003).

32. H. Rigby, C. V. Fernandez, Providing research results to study participants: support versus practice of researchers presenting at the American Society of Hematology annual meeting. *Blood* **106**, 1199–1202 (2005).
33. J. Wade, J. L. Donovan, J. Athene Lane, D. E. Neal, F. C. Hamdy, It's not just what you say, it's also how you say it: Opening the 'black box' of informed consent appointments in randomised controlled trials. *Soc. Sci. Med.* **68**, 2018–2028 (2009).
34. N. Mills, *et al.*, Training recruiters to randomized trials to facilitate recruitment and informed consent by exploring patients' treatment preferences. *Trials* **15**, 323 (2014).
35. A. P. Zwane, *et al.*, Being surveyed can change later behavior and related parameter estimates. *Proc. Natl. Acad. Sci.* **108**, 1821–1826 (2011).

Supporting Appendix

Structured Ethics Appendix Questions: Brief Explanations

1. Policy Equipoise

Is there policy equipoise? That is, is there uncertainty regarding participants' net benefits from each arm of the study relative to the other arms and to the best possible policy to which participants could have access? If not, ethical randomization requires two conditions related to scarcity: (1) Was there scarcity, i.e., did the inclusion of multiple arms change the expected aggregate value of the programs delivered? (2) Do all ex-ante identifiable participants have equal moral or legal claims to the scarce programs?

If there is no reasonable expectation that one arm of the study produces more benefits to participants than any other arm or than the best possible alternative policy, then randomization is ethically unproblematic. If not, then excluding some participants from the superior treatment arm can only be justified by scarcity. Scarcity conditions are two-fold: (1) resources are not sufficient, given constraints, to include all participants in the superior treatment arm; (2) no ex-ante identifiable participants are excluded from the superior arm and have a greater claim to those resources than any participant assigned to the superior arm. See MacKay (2018, 2020) for more complete discussions of policy equipoise.

2. Role of researchers with respect to implementation

Are researchers “active” researchers, i.e. did the researchers have direct decision making power over whether and how to implement the program? If YES, what was the disclosure to participants and informed consent process for participation in the program? Providing IRB approval details may be sufficient but further clarification of any important issues should be discussed here. If NO, i.e., implementation was separate, explain the separation.

A researcher should be considered “active” if, for example, the implementing staff are employed by an institution at which the PI is employed, and the staff report either directly or indirectly to the PI at this institution with regard to this project. Or if researchers control funding for implementation, or have direct decision-making power over key implementation decisions.

Some key factors that help illuminate whether the researchers are “active” or not (here “researchers” are defined as the PIs and the staff that report directly or indirectly to the PIs): Did researchers directly provide any of the interventions, or parts thereof, to participants? Did researchers interact directly with participants and implicitly endorse one or more of the interventions?

3. Potential harms to participants or nonparticipants from the interventions or policies

Does the intervention, policy or product being studied pose potential harm to participants or non-participants? Related, are participants or likely affected non-participants particularly vulnerable? Also related, are participants' access to future services or policies changed because of participation in the study? If yes to any of the above, what is being done to mitigate such risks?

It may be important to consider whether the researchers are “active” (see above) or not for this discussion. If the researchers are “active”, then they are responsible for the potential harms, and thus a robust discussion is appropriate. If the researchers are not “active”, then while they may not be responsible for potential harms, a discussion of this would be appropriate here.

There will almost always be some potential harms, if nothing else because of complementary investments such as time that participants in an intervention necessarily redirect from one activity to another.

Quantifying these risks and complementary investments may be difficult ex-ante, but a discussion of what they are here would help the reader assess their likely importance relative to the potential benefits of the tested intervention. Also note that measuring any harms ex-post may be the exact reason for the study, particularly when the intervention is common.

If risks to nonparticipants exist, discuss the mechanisms through which the risk arises from the study and provide an estimate of the magnitude of the risk and the probability of harm.

4. Potential harms to research participants or research staff from data collection (e.g., surveying, privacy, data management) or research protocols (e.g., random assignment)

Are data collection and/or research procedures adherent to privacy, confidentiality, risk-management, and informed consent protocols with regard to human subjects? Are they respectful of community norms, e.g., community consent not merely individual consent, when appropriate? Are there potential harms to research staff from conducting the data collection that are beyond “normal” risks?

Example of sub-questions to consider as part of the broad question: Are there any risks that could ensue because of the data collection process or storage, e.g. discomfort to being asked certain questions or breach of confidentiality? If so, what are the mitigation strategies? Are there costs to the participant for the data collection process, such as their time, and if so, what is the strategy or rationale for offsetting this cost?

Because these are all issues covered by most IRB processes, a sufficient explanation for a “yes” response may be to provide the IRB approval numbers for all IRBs that have approved the project. However, if there are particular issues that are important to discuss, please do so here.

Harms to research staff could include, e.g., exposure to political violence, exposure to unusual levels of a communicable disease, mistrust due to lack of perceived lack of community consent, or emotional wellbeing from surveying about difficult subject matters. This would not include, e.g., traffic accidents.

5. Financial and reputational conflicts of interest

Do any of the researchers have financial conflicts of interest with regard to the results of the research? Do any of the researchers have potential reputational conflicts of interest?

We define financial conflicts of interest as that used by the researcher’s institutional (e.g., their university) guidelines. We define a reputational conflict of interest as one in which prior writing or advocacy could be contradicted by specific results pursued in this study, and such contradiction would pose reputational risks to the author.

6. Intellectual freedom

Were there any contractual limitations on the ability of the researchers to report the results of the study? If so, what were those restrictions, and who were they from?

This could include, for example, approval of release of the paper and restrictions on data release, but does not include things such as a “comment period” during which interested parties have a right to review and provide comments prior to release but not to control the outputs of the study.

7. Feedback to participants or communities

Is there a plan for providing feedback on research results to participants or communities? If yes, what is the plan? If not, why not?

Engaging in post-study feedback is a way of acknowledging the agency of participants and communities, and is thus a desired practice. However, it may be impractical due to costs, timing, challenges communicating the results, or potential harms if such communication may itself change behavior in undesirable ways.

8. Foreseeable misuse of research results

Is there a foreseeable and plausible risk that the results of the research will be misused and/or deliberately misinterpreted by interested parties to the detriment of other interested parties? If yes, please explain any efforts to mitigate such risk.

In settings with strong imbalances of power between interested parties, there may be foreseeable risks that a powerful party could use deliberately selected research findings to their advantage and to the harm of participants or non-participants, including for general public policy. For example, if the research might reveal the vulnerability of some that can be exploited for the gain of the more powerful party, what steps does the researcher plan to mitigate this risk?

9. Other Ethics Issues to Discuss

Are there any other issues to discuss?